Automatic detection of preclinical neurodegeneration

Presymptomatic Huntington disease

S. Klöppel, MD
C. Chu, MSc
G.C. Tan, BSc
B. Draganski, MD
H. Johnson, PhD
J.S. Paulsen, PhD
W. Kienzle, PhD
S.J. Tabrizi, MD
J. Ashburner, PhD
R.S.J. Frackowiak, MD
PREDICT-HD
    Investigators of the
    Huntington Study
    Group*

Address correspondence and reprint requests to Dr. Stefan Klöppel, Department of Psychiatry, University Clinic Freiburg, Hauptstr. 5, Freiburg, Germany
stefan.kloeppel@uniklinik-freiburg.de

ABSTRACT

Background: Treatment of neurodegenerative diseases is likely to be most beneficial in the very early, possibly preclinical stages of degeneration. We explored the usefulness of fully automatic structural MRI classification methods for detecting subtle degenerative change. The availability of a definitive genetic test for Huntington disease (HD) provides an excellent metric for judging the performance of such methods in gene mutation carriers who are free of symptoms.

Methods: Using the gray matter segment of MRI scans, this study explored the usefulness of a multivariate support vector machine to automatically identify presymptomatic HD gene mutation carriers (PSCs) in the absence of any a priori information. A multicenter data set of 96 PSCs and 95 age- and sex-matched controls was studied. The PSC group was subclassified into three groups based on time from predicted clinical onset, an estimate that is a function of DNA mutation size and age.

Results: Subjects with at least a 33% chance of developing unequivocal signs of HD in 5 years were correctly assigned to the PSC group 69% of the time. Accuracy improved to 83% when regions affected by the disease were selected a priori for analysis. Performance was at chance when the probability of developing symptoms in 5 years was less than 10%.

Conclusions: Presymptomatic Huntington disease gene mutation carriers close to estimated diagnostic onset were successfully separated from controls on the basis of single anatomic scans, without additional a priori information. Prior information is required to allow separation when degenerative changes are either subtle or variable. *Neurology*® 2009;72:426–431

GLOSSARY

AD = Alzheimer disease; **CI** = confidence interval; **DWI** = diffusion-weighted imaging; **FWE** = family-wise error; **HD** = Huntington disease; **PSC** = presymptomatic Huntington disease gene mutation carrier; **ROI** = region of interest; **SVM** = support vector machine; **VBM** = voxel-based morphometry.

Group studies in familial Alzheimer disease (AD)[1] or Huntington disease (HD)[2] have shown substantial neurodegeneration before the onset of typical clinical symptoms. Preclinical degeneration, detectable by standard MRI scans, implies a substantial functional reserve, which indicates that therapeutic attempts to limit degenerative damage are disadvantaged when delayed until a disease is manifest clinically. Consequently, there is a principled need for accurate, early, preclinical diagnosis. Time-efficient methods, applicable with little or no expert knowledge, would be advantageous for screening large numbers of subjects.

Machine-learning techniques meet these requirements. They are fully automatic and have been used to successfully separate magnetic resonance (MR) images on the basis of group characteristics such as sex, or presence/absence of disease.[3-11] Methods such as support vector

Table Demographic information on whole group and on subgroups separated by estimated probability of developing symptoms in 5 years

	Full group		Near group		Mid group		Far group	
	HD	Controls	HD	Controls	HD	Controls	HD	Controls
n	96	95	32	32	32	32	32	32
Sex, F/M	68/28	66/29	21/11	21/11	26/6	24/8	21/11	24/8
Age at MRI scan, mean (range), y	41.8 (25–70)	44.6 (20–63)	46.1 (28–70)	47.2 (30–63)	44.2 (28–70)	44.6 (20–63)	35.1 (25–54)	36.45 (19–52)
CAG, mean (range)	42.9 (39–61)	NA	44.69 (40–61)	NA	42.46 (39–48)	NA	41.47 (39–45)	NA
Probability of symptom onset in next 5 y, %	25.1 (0.7–76.1)	NA	49.29 (35–76)	NA	22.1 (10–33)	NA	3.87 (0.7–9.5)	NA

HD = Huntington disease; NA = not applicable.

machines (SVMs)[12] require well-defined training images from which they learn to separate diagnostic categories. The application of automatic classification methods is often limited by lack of a diagnostic gold standard for validation.

Presymptomatic HD is an important model for study of the earliest stages of neurodegeneration and atrophy because this autosomal dominant disorder has complete penetrance and results from an expanded CAG trinucleotide repeat in the huntingtin gene that is readily detectable in the blood.[13]

Because machine-learning techniques can potentially be used in large, multicenter treatment trials,[14,15] we sought to explore SVM performance on images from several centers. Encouraging SVM performance with HD will support the strategy of using a similar approach to identify a preclinical phase in other neurodegenerative disorders, such as AD.

METHODS Subjects. A cohort of 96 PSCs and 95 control subjects enrolled in the PREDICT-HD study[15] were included. PREDICT-HD is an international multicenter study to discover biologic and refined clinical predictors of disease progression in PSCs. Inclusion criteria for PSCs included at least 39 CAG repeats in the HD gene, whereas controls had fewer than 30 repeats. Exclusion criteria for both PSCs and controls included evidence of unstable illness, alcohol or drug abuse, a history of special educational needs, and a history of other CNS diseases or events.[15] All T1-weighted anatomic brain MRI scans were checked for artifacts using a semiautomatic quality control procedure at the time of acquisition.

PSCs were stratified by their estimated time to clinical manifestation based on age and CAG repeat length (algorithm available at http://www.cmmt.ubc.ca/clinical/hayden).[16] This is a robust model for age of disease diagnosis based on data from almost 3,000 gene carriers. As in previous work on the PREDICT-HD data,[15] we used the algorithm to estimate the probability of developing unequivocal signs of HD in the next 5 years. PSCs were classified into three equally sized subgroups with 1) less than 10%, 2) 10% to 33%, and 3) more than 33% probability of clinical manifestation in 5 years. Controls were matched to each PSC subgroup to achieve the best possible age match; a control subject could serve in more than one group. See the table for full details. The study was performed according to the Declaration of Helsinki and was approved by the ethics review boards of each participating center. All subjects gave written informed consent.

MRI and processing. T1-weighted MRI scans were acquired using a three-dimensional volumetric spoiled gradient echo series on 1.5-tesla scanners (echo time 3 msec, repetition time 18 msec, flip angle 20°, field of view 240 mm, 124 slices at 1.5 mm thickness, matrix size 256 × 192). Because data were acquired from several centers, different hardware was used so small deviations from these sequence parameters were allowed. Where available, phased arrays were preferred over quadrature head coils because of increased signal-to-noise ratio. There was no systematic difference in scanning parameters between groups because participating centers acquired data from PSCs and controls using the same setup. Images were first segmented into gray matter, white matter, and CSF using statistical parametric mapping software, SPM5 (Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, UK; http://www.fil.ion.ucl.ac.uk/spm). Then, gray matter segments were normalized to the population templates generated from all study images using a diffeomorphic registration algorithm.[17] A separate "modulation" step[18] was used to ensure that the overall amount of each tissue class remained constant after normalization. After these steps, the value of a voxel reflects the local gray matter volume.

To evaluate and illustrate the extent of differences in regional gray matter volume between controls and the three PSC subgroups, we performed an analysis using voxel-based morphometry (VBM)[18,19] and applied an exploratory threshold at $p = 0.001$ (uncorrected for multiple comparisons). After preprocessing as above, we smoothed with an 8-mm gaussian kernel and contrasted PSC groups with controls to identify areas with gray matter atrophy. The T scores at the voxels showing the most significant differences in each contrast are reported.

Support vector classification. In what follows, we provide an intuitive understanding of linear SVMs and how they are implemented in the current work. A more technical account of this method can be found in the e-Methods on the *Neurology*® Web site at www.neurology.org, in textbooks,[12,20] or in our previous work.[6,7]

Figure 1   Illustration of the concept used in support vector machines
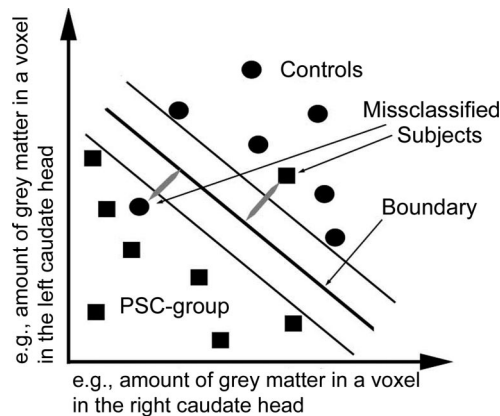
The algorithm tries to find a (multidimensional) boundary that maximizes the distance between presymptomatic Huntington disease gene mutation carriers (squares) and controls (circles). The figure reduces the problem to two dimensions or voxels for the purpose of illustration only. Cases wrongly classified (with gray connectors) are penalized with a specific weight (see Methods for details).

For the present work, we focused on the gray matter segment, because any neurodegenerative process is likely to manifest in that tissue class. We used an off-the-shelf linear SVM (http://www.csie.ntu.edu.tw/˜cjlin/libsvm/). In a first step, all but one scan from PSCs and controls is used to train an SVM. During this training process, all image characteristics (i.e., the gray matter volume in a brain region as reflected by the value of a voxel) are used to define a boundary that separates diagnostic groups. Figure 1 illustrates the principles of SVM in two dimensions (i.e., each subject has two image characteristics or voxels). In practice, there would be several thousand voxels (features) in an image, each of which forms a separate dimension. During this training process, those subjects that are most difficult to separate are used to define the boundary between the diagnostic groups. Sometimes there is too much overlap between the groups, in which case higher accuracy can be achieved by allowing some of the training data to fall on the wrong side of the boundary. A parameter C is used to control how much misclassification of the training data is allowable.

The next step is to ensure that this boundary is useful to correctly separate new data. These new data do not contribute to the definition of the classification boundary. In the clinical setting, this new data could come from a patient to be diagnosed. In our implementation, we used a further round of training and testing to optimize the C parameter (called a three-way split validation). We report the average accuracy, i.e., what percentage of scans left out of the training set were assigned to their correct group. This percentage can be converted into a p value by assumption of a binomial distribution with a chance probability of correct classification of 0.5.

Another way to check whether the classification boundary relies on meaningful information is to localize the pattern of voxels characterizing differences between groups (figure 2 and e-Methods).

Region of interest approach. Optimally, classification methods should be capable of detecting preclinical degeneration without additional disease-specific information. Such information will often be unavailable because symptoms may be subtle and nonspecific or absent, or the earliest site of pathology is unknown. In an additional analysis, we explored whether an improvement of classification accuracy accrues with addition of prior information about regions known to be affected by the disease. HD, like other neurodegenerative diseases, does not affect all brain areas to a similar extent. We used a group comparison between normal and PSC scans with VBM to generate a weighted map of areas involved by preclinical neurodegeneration. Meaningful classification by SVMs has to generalize to new images. If regions involved in a disease process had been identified from the same data set later used for classification, any differences could be specific only to the scans of that data set, but not to the underlying disease process in general. Therefore, to ensure generalization of results, a separate data set of PSC and control scans was acquired using three-dimensional structural magnetic resonance images from 42 PSCs and control subjects to define the regions of interest (see table e-1 for demographic details). We used a 1.5-tesla Siemens Sonata scanner (T1-weighted MDEFT sequence, 176 slices at 1-mm thickness, sagittal, phase encoding in anterior/posterior, field of view 224 × 256 mm², matrix 224 × 256, repetition time 20.66 msec, echo time 8.42

Figure 2   Brain regions most relevant for group separation
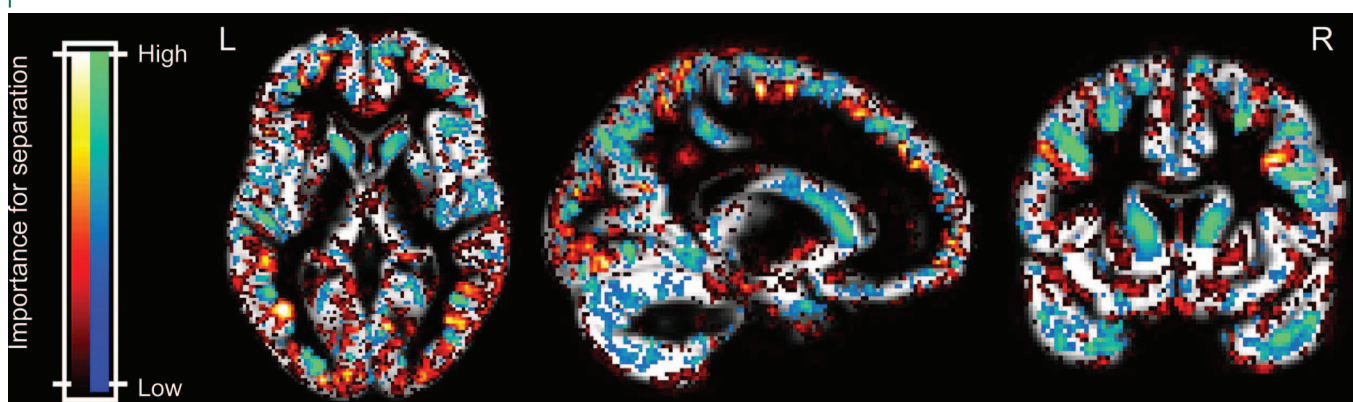
Voxels most relevant for the separation of Huntington disease mutation carriers and controls in the group closest to clinical presentation. A dual color bar is used. Blue to green areas indicate higher gray matter density, increasing the likelihood of classification as normal. Red and yellow show regions where higher gray matter density characterizes asymptomatic Huntington disease mutation carriers. See main text for interpretation. Results are overlaid on the mean gray matter compartment from all subjects.

msec, inversion time 640 msec, flip angle 25 degrees, fat saturation, bandwidth 178 Hz/pixel).[21]

We used the t value at each voxel to evaluate its involvement in the disease process. To reduce the number of voxels, we restricted this analysis only to those voxels surviving correction for multiple comparisons across whole brain using a FWE correction as implemented in SPM5. To assess the performance of such a region of interest (ROI) definition process, we also generated a weighted image with a more liberal threshold of $p = 0.01$ (uncorrected; see figure e-1 for resulting T-maps).

Previous imaging studies have shown that the striatum is most affected and that it atrophies early in HD.[2] The degree of atrophy is comparable to the early degeneration of the hippocampus in AD. In the extreme case, we selected our region for categorization as that with the coordinates of the single voxel showing the greatest atrophy (indexed by the highest VBM T score) in the VBM comparison of the separate groups of PSCs and control. To include maximum a priori information from this comparison, we applied the same amount of gaussian smoothing to the classifier images as applied to the VBM ROI-defining image set before extracting the voxel value.

RESULTS Categorization accuracy depended greatly on estimated time to disease presentation. Subjects with at least a 33% chance of developing unequivocal signs of HD in 5 years were correctly assigned, with no a priori information, to the PSC group 69% ($p = 0.002$) of the time. Best performance (82.8%; $p < 0.001$) was obtained with the weighted VBM voxel procedure. Classification accuracy for the PSC group furthest from clinical onset was at chance. Figure 2 illustrates that the striatum was critical for a separation of controls from preclinical HD subjects. The distribution of blue and green colors also indicates that in regions including the insula and parts of the parietal cortex, reduced gray matter was indicative of PSC status. The effect of different levels of a priori information for each of the groups is demonstrated in figure 3. Table e-2 summarizes all the results and provides specificity and sensitivity values together with confidence intervals (CIs).

Even at an exploratory threshold, no significant VBM gray matter differences were found between controls and the PSC group with a less than 10% probability of developing symptoms in 5 years. Subjects closer to estimated onset and subjects from all subgroups combined showed the expected gray matter loss in the striatum compared with controls (figure e-2). In these group comparisons, the maximal difference was located in striatum (combined group T = 7.67; group closest T = 7.98; middle group T = 7.39).

DISCUSSION We sought to characterize the ability of a fully automatic image classification method to separate structural MRI brain scans of HD gene carriers in the presymptomatic phase from those of con-

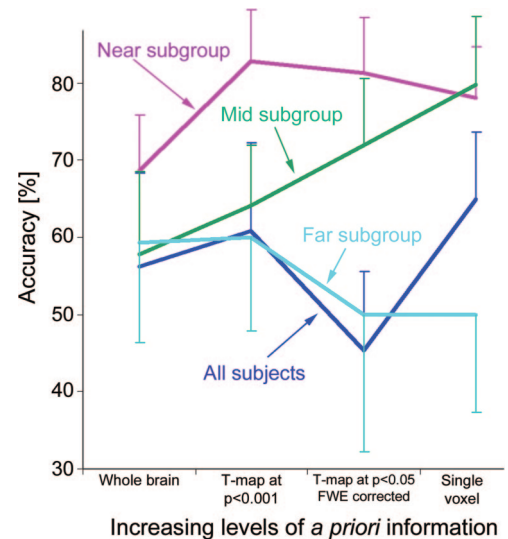Figure 3   Classifications accuracy for different levels of a priori information

Illustration of accuracy in each subgroup and with all subjects pooled together for different levels of a priori information. Error bars indicate 95% confidence intervals (calculated by the efficient-score method[29]; http://faculty.vassar.edu/lowry/clin1.html). One-sided error bars are used for illustration purposes. FWE = family-wise error.

trols. Subjects with a more than 33% probability of clinical diagnosis of HD within 5 years were correctly separated from controls 69% of times without any a priori regional weighting. Although this accuracy is clearly above chance (see CIs in figure 3), it is nowhere near perfect. It is interesting that whole brain classification accuracy—this study—falls substantially below the 82% correct classification achieved in an earlier study using an SVM on diffusion-weighted imaging (DWI) data less readily available in clinical practice than T1-weighted data.[6] Subjects in the DWI study were unrelated to those of this one and as a group were estimated to be on average 19 years from clinical presentation. Although CIs will overlap, the suggestion is that diffusion imaging is better at classifying HD images. This conclusion is at odds with results from (univariate) VBM studies that show highly significant differences between PSC and control group T1-weighted images[2] that are larger than those obtained using DWI.[22,23] The differences in acquisition time (10 minutes for a T1 compared with 22 minutes [12 minutes without cardiac gating] for a DWI sequence) and the fact that the study reported here used a multicenter data set are two likely explanations for this apparent disagreement.

As expected, classification accuracy improved for PSC subjects closest to estimated symptom onset. The best performance was achieved when brain areas used for classification were limited to regions identified by VBM as affected in the PSC group. In gen-

eral, a multivariate method that includes information from various brain areas should show favorable performance when more voxels (reflecting the volume of more brain regions) yield relatively more signal than noise. Figure 2 illustrates that group separation relies heavily on voxels within the caudate nucleus and particularly its head. Reduced gray matter reflected PSC status in insula and parietal cortex also; findings well in line with previous imaging studies.[2,24] The figure also displays cortical voxels scattered throughout the brain, without a regionally specific pattern. These scattered voxels constitute a source of "noise," which explains the superior performance of classification using the caudate alone, a procedure equivalent to minimizing noise.

Figure 3 illustrates the benefit of various levels of a priori information, which becomes most obvious for subjects in the middle group but also when all subjects are combined. In contrast, no meaningful classification accuracy was achieved in subjects far from estimated clinical onset, no matter how much a priori information was used.

VBM-derived prior information from an independent set of images served two purposes. We avoided overoptimistic claims and any circular logic about result generalization, which would have arisen had we created VBM-weighted images from the images that were also classified with SVM. VBM analysis also created a specific weighted group image that characterized the preclinical HD phase. The creation of similarly informative images could have been achieved using atlas-based masks of putamen and caudate. The approach we present here is more flexible. It allows the creation of disease-specific weighted images when disease distribution does not respect anatomic boundaries or is more widespread. A further advantage of our approach is that each voxel obtains a specific weighting. In contrast, anatomically based masks are normally binary and hence less specific. As expected, no improvement of classification was achieved when VBM derived T-maps were binarized (data not shown). Relatively labor-intensive manual outlining methods, often used in HD,[25] would be less suitable for screening than the one presented here. A study comparing both approaches in early HD[26] found that both methods reliably showed expected degeneration, but VBM detected additional changes in brain regions not selected a priori.

Performance was at chance level when we attempted to separate the subgroup far from clinical presentation from matched controls. Depending on the individual number of CAG repeats and age, subjects in this group were an estimated 20 years or more from developing signs of disease. It is a matter of debate when striatal degeneration starts. A large-scale study based on striatal volume change in PSC[27] illustrates that decline of striatal volume is very subtle in subjects with more than 20 years to estimated onset but becomes substantially steeper around 15 years beforehand. VBM analysis confirms that structural changes were either absent or too subtle in the group farthest from onset to be detected in a group-level VBM analysis. In contrast, bilateral striatal gray matter loss found in the other subgroups confirms previous work using VBM.[2]

Classification performance was far from perfect. There is a wide range of techniques for extracting image characteristics to feed into various classification methods.[3,9,11,28] The purpose of our study was to test gray matter–based SVM classification successfully applied to patients with mild to moderate AD on preclinical HD.[7] The study in AD demonstrated the utility when cases were at a point where clinical signs were significant and disease-related atrophy was significant. Here we use genetic information not only to recruit individuals before the manifestation of any clinical deterioration, but also to estimate years to onset of disease and thus make use of the technique to detect the earliest and most subtle degenerative change in the brain. Both studies used data acquired at multiple imaging centers. Although this has to be shown for each disease, our work suggests that data can be exchanged between centers. If this proves true for other diseases, it would make excessive data acquisitions unnecessary and would facilitate the application to rarer neurodegenerative disorders.

Our results show that fully automatic detection of preclinical degeneration is possible so that identified subjects could become candidates for longitudinal follow-up in clinical trials, possibly many years before clinical presentation.[27] It will be another topic of future studies to test whether multivariate classification methods such as those presented here can play a part in the detection of longitudinal changes alongside currently used, well-established imaging, cognitive, and behavioral changes.[27]

AUTHOR CONTRIBUTIONS

S.K., C.C., G.C.T., B.D., S.J.T., J.A., and R.S.J.F. planned and designed the study. J.S.P. and H.J. provided quality control data. S.K., C.C., W.K., and J.A. were involved in the analyses. All authors contributed to the manuscript and approved the final version.

REFERENCES

1. Fox NC, Crum WR, Scahill RI, Stevens JM, Janssen JC, Rossor MN. Imaging of onset and progression of Alzheimer's disease with voxel-compression mapping of serial magnetic resonance images. Lancet 2001;358:201–205.

2. Thieben MJ, Duggins AJ, Good CD, et al. The distribution of structural neuropathology in pre-clinical Huntington's disease. Brain 2002;125:1815–1828.

3. Davatzikos C, Fan Y, Wu X, Shen D, Resnick SM. Detection of prodromal Alzheimer's disease via pattern classification of MRI. Neurobiol Aging 2008;29:514–523.

4. Fan Y, Shen D, Davatzikos C. Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM. Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv 2005;8:1–8.

5. Kawasaki Y, Suzuki M, Kherif F, et al. Multivariate voxel-based morphometry successfully differentiates schizophrenia patients from healthy controls. Neuroimage 2007;34:235–242.

6. Kloppel S, Draganski B, Golding CV, et al. White matter connections reflect changes in voluntary-guided saccades in pre-symptomatic Huntington's disease. Brain 2008;131:196–204.

7. Kloppel S, Stonnington CM, Chu C, et al. Automatic classification of MR scans in Alzheimer's disease. Brain 2008; 131:681–689.

8. Lao Z, Shen D, Xue Z, Karacali B, Resnick SM, Davatzikos C. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. Neuroimage 2004;21:46–57.

9. Teipel SJ, Born C, Ewers M, et al. Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. Neuroimage 2007;38:13–24.

10. Teipel SJ, Stahl R, Dietrich O, et al. Multivariate network analysis of fiber tract integrity in Alzheimer's disease. Neuroimage 2007;34:985–995.

11. Vemuri P, Gunter JL, Senjem ML, et al. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. Neuroimage 2008;39:1186–1197.

12. Vapnik V. Statistical Learning Theory. New York: Wiley Interscience; 1998.

13. Huntington Study Group. Unified Huntington's Disease Rating Scale: reliability and consistency. Mov Disord 1996;11:136–142.

14. Mueller SG, Weiner MW, Thal LJ, et al. The Alzheimer's disease neuroimaging initiative. Neuroimaging Clin North Am 2005;15:869–877, xi–xii.

15. Paulsen JS, Hayden M, Stout JC, et al. Preparing for preventive clinical trials: the Predict-HD study. Arch Neurol 2006;63:883–890.

16. Langbehn DR, Brinkman RR, Falush D, Paulsen JS, Hayden MR. A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. Clin Genet 2004;65:267–277.

17. Ashburner J. A fast diffeomorphic image registration algorithm. Neuroimage 2007;38:95–113.

18. Ashburner J, Friston KJ. Voxel-based morphometry: the methods. Neuroimage 2000;11:805–821.

19. Wright IC, McGuire PK, Poline JB, et al. A voxel-based method for the statistical analysis of gray and white matter density applied to schizophrenia. Neuroimage 1995;2:244–252.

20. Bishop C. Pattern Recognition and Machine Learning. New York: Springer; 2006.

21. Deichmann R, Schwarzbauer C, Turner R. Optimisation of the 3D MDEFT sequence for anatomical brain imaging: technical implications at 1.5 and 3 T. Neuroimage 2004;21:757–767.

22. Reading SA, Yassa MA, Bakker A, et al. Regional white matter change in pre-symptomatic Huntington's disease: a diffusion tensor imaging study. Psychiatry Res 2005;140:55–62.

23. Rosas HD, Tuch DS, Hevelone ND, et al. Diffusion tensor imaging in presymptomatic and early Huntington's disease: selective white matter pathology and its relationship to clinical measures. Mov Disord 2006;21:1317–1325.

24. Rosas HD, Hevelone ND, Zaleta AK, Greve DN, Salat DH, Fischl B. Regional cortical thinning in preclinical Huntington disease and its relationship to cognition. Neurology 2005;65:745–747.

25. Aylward EH, Sparks BF, Field KM, et al. Onset and rate of striatal atrophy in preclinical Huntington disease. Neurology 2004;63:66–72.

26. Douaud G, Gaura V, Ribeiro MJ, et al. Distribution of grey matter atrophy in Huntington's disease patients: a combined ROI-based and voxel-based morphometric study. Neuroimage 2006;32:1562–1575.

27. Paulsen JS, Langbehn DR, Stout JC, et al. Detection of Huntington's disease decades before diagnosis: the Predict HD study. J Neurol Neurosurg Psychiatry 2008;79:874–880.

28. Fan Y, Batmanghelich N, Clark CM, Davatzikos C. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. Neuroimage 2008;39: 1731–1743.

29. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med 1998;17:857–872.

No Charge for Color Figures

Neurology® is committed to presenting data in the most descriptive way for the benefit of our readers. To make possible the publication of a greater number of color figures, we have elimated our color figure charges to authors.